ON THE CONVERGENCE OF THE ITERATIVE "PSEUDO LIKELIHOOD" MAXIMIZATION ALGORITHM

Cédric Herzet

IRISA-INRIA, 35042 Rennes cedex, France

ABSTRACT

We consider a powerful iterative inference algorithm which has recently appeared in the literature, see e.g. [1, 2, 3, 4, 5]. In this paper, we refer to this algorithm as iterative "pseudo likelihood" maximization (IPLM) algorithm. We give a connection between this algorithm and the problem of Bethe free energy minimization and prove several important results concerning its fixed points and its convergence properties.

***Index Terms*—** MAP estimation, iterative methods, convergence of numerical methods.

1. INTRODUCTION

In this paper, we consider the problem of inferring the value of an unknown parameter Θ from an observation vector $\mathbf{Y}$, when $\mathbf{Y}$ also depends on a "nuisance" parameter vector $\mathbf{X} = [\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_N]$. In such a scenario, a direct computation of the maximum a posteriori (MAP) estimate often turns out to be a complex task. In order to circumvent this problem, powerful numerical methods, enabling to iteratively compute the MAP solution, have been proposed in the literature. For example, the expectation-maximization (EM) algorithm [6] or the family of gradient methods [7] are instances of such algorithms. More recently, iterative estimation methods based on factor graphs (FGs) and the belief-propagation (BP) algorithm [8] have appeared in the literature, see e.g. [1, 2, 3]. Although slightly different in their implementation, these methods have the common feature of computing a sequence of estimates $\{\theta^{(n)}\}_{n=0}^{\infty}$ by increasing at each iteration a "pseudo" likelihood function (PLF); the latter likelihood function being built by considering standard BP messages as a priori information on the nuisance parameters. In the remainder of this paper, we will therefore refer to this kind of algorithm as iterative "pseudo likelihood" maximization (IPLM) algorithm.

In [3], the authors proposed to maximize the PLF by means of the EM algorithm. Considering this particular implementation in *cycle-free* FGs, they showed that if only one EM iteration is performed, one recovers the standard implementation of the EM algorithm, proving as a by-product that the fixed points of this particular IPLM algorithm must be stationary point of the true likelihood function. This conclusion was later shown to be valid irrespective of the method used to maximize the PLF in two parallel works [4, 5]: in [4] this result was shown in the particular context of synchronization problems whereas general FGs were considered in [5]. In [5], the author also provides a result for *cyclic* FG, but the proposed characterization does not enable a simple interpretation of the nature of the fixed points. In this contribution, we prove several important properties of the IPLM algorithm for general FGs by placing the MAP estimation problem in the more general framework of Bethe free energy minimization. In particular, we show that *i)* the fixed points of the IPLM algorithm must be stationary points of the *Bethe free energy* of the system [9]; *ii)* any fixed point of the IPLM is also a fixed point of the EM algorithm; *iii)* we formulate necessary and sufficient conditions for local convergence of the IPLM algorithm. As a corollary, we show that the IPLM algorithm never converges to maxima of the Bethe free energy.

2. BETHE FREE ENERGY MINIMIZATION

In order to generalize the properties of the IPLM algorithm to general FGs, we will consider the general problem of Bethe free energy minimization. Assume that

$$p_{\Theta,\mathbf{X},\mathbf{Y}}(\theta,\mathbf{x},\mathbf{y}) = \prod_{a=1}^{M} \Psi_{\mathbf{X}_{V_a},\Theta}(\mathbf{x}_{V_a},\theta), \tag{1}$$

where $V_a \subset \{1, 2, \ldots, N\}$ and $\mathbf{X}_{V_a}$ is a vector made up of the elements of $\mathbf{X}$ whose index is in V_a. Let us consider the FG associated to (1) where Θ is *not* a variable node but simply a parameter of the factor nodes (i.e. only $X_1, \ldots, X_N$ are variable nodes in the FG). If the FG is cycle free, it is well-known [9] that

$$\log p_{\Theta,\mathbf{Y}}(\theta,\mathbf{y}) = -G_{\Theta,B_a(\mathbf{x}_{V_a}),B_i(x_i)}\left(\theta, b_a^{\star}(\mathbf{x}_{V_a}), b_i^{\star}(x_i)\right),$$

where $G_{\Theta,B_a(\mathbf{x}_{V_a}),B_i(x_i)}(\cdot)$ is the Bethe free energy associated to the FG and $b_a^{\star}(\mathbf{x}_{V_a})$ (resp. $b_i^{\star}(x_i)$) are the beliefs computed by the BP algorithm [8] at the factor (resp. variable) nodes of the FG, i.e.

$$b_a^{\star}(\mathbf{x}_{V_a}) = \gamma_a^{-1} \Psi_{\mathbf{X}_{V_a},\Theta}(\mathbf{x}_{V_a},\theta) \prod_{i \in V_a} \mathbf{m}_{i\to a}(x_i,\theta) \tag{2}$$

$$b_i^{\star}(x_i) = \gamma_i^{-1} \prod_{a \in P_i} \mathbf{m}_{a\to i}(x_i,\theta), \tag{3}$$

where $\mathbf{m}_{a\to i}(x_i,\theta)$ and $\mathbf{m}_{i\to a}(x_i,\theta)$ are the messages computed by the BP algorithm, P_i is the set of factor nodes connected to variable node i, and γ_a, γ_i are normalization factors. In the cycle free case, we see that minimizing the Bethe free energy with respect to Θ is equivalent to compute the MAP estimate. In the sequel, we will therefore focus on the following more general problem:

$$\theta^\star = \arg\max_\theta L_\Theta(\theta), \tag{4}$$

where

$$L_\Theta(\theta) = -G_{\Theta,B_a(\mathbf{x}_{V_a}),B_i(x_i)}\left(\theta, b_a^\star(\mathbf{x}_{V_a}), b_i^\star(x_i)\right). \tag{5}$$

Problem (4) has the following interpretation: when the FG associated to (1) is cycle free, $\theta^\star$ is the MAP solution; otherwise it is an approximation the MAP solution.

3. THE IPLM ALGORITHM: PROPERTIES

In this section, we will derive several properties of the IPLM algorithm. We proceed in two steps. We first define the "pseudo" likelihood function (PLF) and emphasize some of its properties. Based on these results, we will then study the convergence properties of the IPLM algorithm.

3.1. PLF: Definition and Properties

Before giving a definition of the PLF, we need to define some notations and concepts.

Definitions: A *region* $\mathcal{R}$ of a FG is defined by a set of factor nodes and the set of *all* variable nodes that are connected to them. A *covering set* Ω is a set of regions such that all factor nodes in the FG are included in one and only one region of the set. A variable node i is said to be a *boundary node* if there exists at least one factor node a such that $a \notin \mathcal{R}$ and $a \in P_i$.

Notations: $V_\mathcal{R}$ (resp. $P_\mathcal{R}$) is the set of index of variable nodes (resp. factor nodes) belonging to region $\mathcal{R}$; $V_\mathcal{R}^B$ is the set of index of the boundary variable nodes belonging to region $\mathcal{R}$.

Let Ω be a covering set of *cycle-free* regions. We define the *pseudo likelihood function* associated to Ω as

$$G^\Omega_{\Theta,\Theta'}(\theta,\theta') \triangleq \sum_{\mathcal{R}\in\Omega} \log \sum_{\mathbf{x}_\mathcal{R}} \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta'), \tag{6}$$

where

$$\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta) \triangleq \prod_{a\in P_\mathcal{R}} \Psi_{\mathbf{X}_{V_a},\Theta}(\mathbf{x}_{V_a},\theta), \tag{7}$$

$$\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta') \triangleq \prod_{i\in V_\mathcal{R}^B} \prod_{a\in P_i\setminus P_R} \mathbf{m}_{a\to i}(x_i,\theta'), \tag{8}$$

i.e. $\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)$ is equal to the product of the factors belonging to $\mathcal{R}$ and $\Phi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)$ is equal to the product of the messages entering the boundary variable nodes of $\mathcal{R}$.

Let us now prove some properties of the region-based free energy which will later prove to be useful in the analysis of the convergence of the IPLM algorithm.

Property 3.1: If $\Theta' = \theta$, we have

$$\nabla_\Theta L_\Theta(\theta) = \nabla_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta). \tag{9}$$

Proof: On the one hand, taking into account the expression of $b_a^\star(\mathbf{x}_{V_a})$ and $b_i^\star(x_i)$ in (2), (3) and applying standard derivation rules, we obtain (see, e.g. [10]):

$$\begin{aligned}\nabla_\Theta L_\Theta(\theta) &= \nabla_\Theta G_{\Theta,B_a(\mathbf{x}_{V_a}),B_i(x_i)}\left(\theta, b_a^\star(\mathbf{x}_{V_a}), b_i^\star(x_i)\right)\\ &= \sum_{a=1}^M \sum_{\mathbf{x}_{V_a}} b_a^\star(\mathbf{x}_{V_a})\, \nabla_\Theta \log \Psi_{\mathbf{X}_{V_a},\Theta}(\mathbf{x}_{V_a},\theta).\end{aligned} \tag{10}$$

On the other hand, taking the derivative of (6) with respect to Θ and using the fact that $\nabla_\Theta \log f_\Theta = \frac{\nabla_\Theta f_\Theta}{f_\Theta}$, we get

$$\nabla_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta')$$

$$= \sum_{\mathcal{R}\in\Omega}\sum_{\mathbf{x}_\mathcal{R}} \frac{\nabla_\Theta\big(\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta')\big)}{\sum_{\mathbf{x}_\mathcal{R}} \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta')}, \tag{11}$$

$$= \sum_{\mathcal{R}\in\Omega}\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta), \tag{12}$$

$$= \sum_{\mathcal{R}\in\Omega}\sum_{a\in P_\mathcal{R}}\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log \Psi_{\mathbf{X}_{V_a},\Theta}(\mathbf{x}_{V_a},\theta), \tag{13}$$

where

$$b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta') \triangleq \frac{\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta')}{\sum_{\mathbf{x}_\mathcal{R}} \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta')}. \tag{14}$$

Now, since Ω is a *covering* set of regions, we have that $\sum_{\mathcal{R}\in\Omega}\sum_{a\in P_\mathcal{R}} = \sum_{a=1}^M$. Moreover, since the regions are cycle free, it can readily be shown starting from (14) and using the definition of the BP message update rules [8] that

$$\sum_{\mathbf{x}_\mathcal{R}\in\mathcal{X}_\mathcal{R}^{\mathbf{x}_{V_a}}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta) = b_a^\star(\mathbf{x}_{V_a}), \tag{15}$$

where $\mathcal{X}_\mathcal{R}^{\mathbf{x}_{V_a}}$ is the set of possible values for $\mathbf{X}_\mathcal{R}$ when $\mathbf{X}_{V_a} = \mathbf{x}_{V_a}$. Therefore, plugging (15) into (13) and comparing with (10) we get (9). $\square$

This first property of the PLF is very interesting since it states that the Bethe free energy and the PLF have locally the same first order behavior. As we will see in the next section, this

property will reveal to be key in the characterization of the fixed points of the IPLM algorithm.

Property 3.2: The Hessian of $L_\Theta(\theta)$ may be expressed as

$$\nabla^2_\Theta L_\Theta(\theta) = \nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta) + \nabla_{\Theta,\Theta'} G^\Omega_{\Theta,\Theta'}(\theta,\theta), \quad (16)$$

where $\nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta)$ and $\nabla_{\Theta,\Theta'} G^\Omega_{\Theta,\Theta'}(\theta,\theta)$ are defined in (17) and (18).

Proof: Starting from (12), we have

$$\begin{aligned}
&\nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta') \\
&= \sum_{\mathcal{R}\in\Omega} \sum_{\mathbf{x}_\mathcal{R}} \nabla_\Theta b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\, \nabla_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta) \\
&+ \sum_{\mathcal{R}\in\Omega} \sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\, \nabla^2_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta).
\end{aligned} \quad (19)$$

Now, using the definition of $b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')$ and the fact that $\nabla_\Theta \log f_\Theta = \frac{\nabla_\Theta f_\Theta}{f_\Theta}$, we obtain after some calculus

$$\begin{aligned}
&\nabla_\Theta b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta') \\
&= b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\Big(\nabla_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta) \\
&- \sum_{\mathbf{x}'_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}'_\mathcal{R},\theta,\theta') \nabla_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}'_\mathcal{R},\theta)\Big).
\end{aligned} \quad (20)$$

Plugging (20) into (19), we get (17). Proceeding in the same way, we can get similar expressions for $\nabla_{\Theta,\Theta'} G^\Omega_{\Theta,\Theta'}(\theta,\theta)$ and $\nabla^2_\Theta L_\Theta(\theta)$ and prove (16). □

4. CHARACTERIZATION OF THE CONVERGENCE OF THE IPLM ALGORITHM

In this section, we study the properties of the following iterative algorithm:

$$\theta^{(n+1)} = \arg\max_\theta G^\Omega_{\Theta,\Theta'}(\theta,\theta^{(n)}). \quad (21)$$

It is easy to see that the definition of (21) includes the IPLM algorithms considered in [1, 2, 3]. Note also that (21) is equivalent to the "Hybrid-EM algorithm" considered in [5] when none of the nodes follow the so-called "E-log" rule [5].

Let us now prove some interesting properties of (21):

Property 4.1: Let θ_f be a fixed point of (21). Then we have

$$\nabla_\Theta G_{\Theta,B_a(\mathbf{x}_{V_a}),B_i(x_i)}\left(\theta_f, b^\star_a(\mathbf{x}_{V_a}), b^\star_i(x_i)\right) = 0. \quad (22)$$

Proof: If θ_f is a fixed point, then we must have

$$\nabla_\Theta G^\Omega_{\Theta,\Theta'}(\theta_f,\theta_f) = 0. \quad (23)$$

Now, from (9) it also implies

$$\nabla_\Theta G_{\Theta,B_a(\mathbf{x}_{V_a}),B_i(x_i)}\left(\theta_f, b^\star_a(\mathbf{x}_{V_a}), b^\star_i(x_i)\right) = 0. \quad (24)$$

and θ_f is a stationary point of the Bethe free energy. □

We see that property 4.1 gives a nice interpretation of the fixed point of the IPLM algorithm in terms of stationary point of the Bethe free energy. Property 4.1 basically states that any fixed point of the IPLM algorithm must be stationary point of the Bethe free energy. This feature is of course highly desirable since any solution of (4) must also cancel the first derivative of the Bethe free energy. Note that in the particular case of FGs without cycles, the Bethe free energy is equal to $-\log p_{\mathbf{Y},\Theta}(\mathbf{y},\theta)$ and we therefore recover the result previously proved in [4, 5].

The next property relates the fixed points of the IPLM algorithm to those of the EM algorithm:

Property 4.2: Let Γ_G denotes the set of fixed points of (21) and let Γ_{EM} denote the set of fixed points of the EM algorithm[1]. Then, we have

$$\Gamma_G \subseteq \Gamma_{EM}. \quad (25)$$

Proof: We must show that if θ_f is a fixed point of (21) then it is also a fixed point of the EM algorithm. Now, any θ_f which satisfies the two following sufficient conditions [10]:

$$\sum_{\mathcal{R}\in\Omega} \sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta_f,\theta_f) \nabla_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta_f) = 0, \quad (26)$$

$$\sum_{\mathcal{R}\in\Omega} \sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta_f,\theta_f) \nabla^2_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta_f) \prec 0. \quad (27)$$

is a fixed point of the EM algorithm. From properties 3.1, 4.1 and (12), we know that the first condition is fulfilled for any fixed point of (21). Let us show that any fixed point of (21) also satisfies the second one. If θ_f is a fixed point of (21), then

$$\nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta) \prec 0. \quad (28)$$

Now, we have

$$\begin{aligned}
&\nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta) \\
&\succeq \sum_{\mathcal{R}\in\Omega} \sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta_f) \nabla^2_\Theta \log \Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta_f).
\end{aligned} \quad (29)$$

since the last two terms in (17) form a definite positive matrix. As a consequence, we see from (29) that (28) also implies (27). □

Property 4.3: The IPLM algorithm never *locally* converges to minima of $L_\Theta(\theta)$. Moreover, it *locally* converges to a maximum of $L_\Theta(\theta)$, say θ_m, if and only if:

$$\nabla_{\Theta,\Theta'} G_{\Theta,\Theta'}(\theta_m,\theta_m) \succ \nabla^2_\Theta\, G_{\Theta,\Theta'}(\theta_m,\theta_m). \quad (30)$$

[1] If the FG is cycle free, we refer to the standard EM algorithm [6]. When the FG contains cycles, we consider the "approximate EM algorithm" described in [11], i.e., the E-step is performed by applying the BP algorithm on the cyclic FG.

$$\nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta') = \sum_{\mathcal{R}\in\Omega}\Bigg(\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla^2_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)$$
$$+\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\,\nabla_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)$$
$$-\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\,\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\Bigg), \tag{17}$$

$$\nabla_{\Theta,\Theta'} G^\Omega_{\Theta,\Theta'}(\theta,\theta') = \sum_{\mathcal{R}\in\Omega}\Bigg(\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\,\nabla_{\Theta'} \log\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta')$$
$$-\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_\Theta \log\Psi_{\mathbf{X}_\mathcal{R},\Theta}(\mathbf{x}_\mathcal{R},\theta)\,\sum_{\mathbf{x}_\mathcal{R}} b_{\mathbf{X}_\mathcal{R},\Theta,\Theta'}(\mathbf{x}_\mathcal{R},\theta,\theta')\nabla_{\Theta'} \log\Phi_{\mathbf{X}_\mathcal{R},\Theta'}(\mathbf{x}_\mathcal{R},\theta')\Bigg). \tag{18}$$

Proof: In order to prove property 4.3, let us consider the following condition of convergence:

$$-\mathbf{I} \prec \mathbf{R}_G(\theta_f) \prec \mathbf{I}, \tag{31}$$

where **I** is the unitary matrix and (see [12])

$$\mathbf{R}_G(\theta_f) = \left(-\nabla^2_\Theta G_{\Theta,\Theta'}(\theta_f,\theta_f)\right)^{-1}\,\nabla_{\Theta,\Theta'} G_{\Theta,\Theta'}(\theta_f,\theta_f). \tag{32}$$

is the (local) rate of convergence of (21) around θ_f. We will show that (31) is never satisfied for minima whereas it is satisfied for maxima if and only if (30) is satisfied.

Taking into account that $\nabla^2_\Theta\, G_{\Theta,\Theta'}(\theta_f,\theta_f) \prec 0$ for any fixed point and using (32) and (16), (31) may be equivalently rewritten as

$$2\nabla^2_\Theta\, G_{\Theta,\Theta'}(\theta_f,\theta_f) \prec \nabla^2_\Theta L_\Theta(\theta) \prec 0. \tag{33}$$

Based on this expression we can draw the two following conclusions. First, if θ_f is a minimum of $L_\Theta(\theta)$, then the IPLM algorithm locally diverges from θ_f. Indeed, if θ_f corresponds to a minimum, it implies $\nabla^2_\Theta L_\Theta(\theta_f) \succ 0$. Therefore, the second inequality in (33) is violated and the algorithm does not converge to θ_f. On the other hand, if θ_f corresponds to a maximum of $L_\Theta(\theta)$, $\nabla^2_\Theta L_\Theta(\theta_f) \prec 0$ and the second inequality in (33) is always satisfied. The (local) convergence to θ_f is therefore ensured if and only if

$$2\nabla^2_\Theta\, G_{\Theta,\Theta'}(\theta_f,\theta_f) \prec \nabla^2_\Theta\, L_\Theta(\theta_f) \tag{34}$$

which is equivalent to (30) by using (16). □

In property 4.1, we saw that some fixed points of the IPLM algorithm can possibly correspond to maxima of the the Bethe free energy. From property 4.3, we see that even if a maximum of the Bethe free energy is a fixed point of (21), the algorithm will not converge to it. Moreover, property 4.3 provides necessary and sufficient conditions (30) for convergence to the minima of the Bethe free energy. This condition is however difficult to check in practice since it basically requires the complex task of evaluating $\nabla^2_\Theta\, L_\Theta(\theta_f)$ (see (34)). However, it gives an interesting insight of when the IPLM algorithm is likely to converge by looking at the expression of $\nabla^2_\Theta G^\Omega_{\Theta,\Theta'}(\theta,\theta')$ and $\nabla_{\Theta,\Theta'} G^\Omega_{\Theta,\Theta'}(\theta,\theta')$ in (17), (18).

5. REFERENCES

[1] L. Zhang and A. Burr. ”APPA Symbol Timing Recovery Scheme for Turbo-codes”. In *IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC’*, Lisbonne, Portugal, Nov. 2002.

[2] J. Dauwels and H.-A. Loeliger. ”Phase Estimation by Message Passing”. In *IEEE International Conference on Communications, ICC’*, pages 523–527, Paris, France, June 2004.

[3] C. Herzet, V. Ramon, and L. Vandendorpe. ”Turbo-synchronization: a Combined Sum-product and Expectation-Maximization Algorithm Approach”. In *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC’*, pages 191– 195, New-York, USA, June 2005.

[4] C. Herzet. ”Code-aided Synchronization for Digital Burst Communications”. PhD thesis, available at http://www.tele.ucl.ac.be/digicom/herzet/index.php.

[5] J. Dauwels. ”On Graphical Models for Communications and Machine Learning: Algorithms, Bounds,and Analog Implementation”. PhD. thesis, Swiss Federal Institute of Technology Zurich, 2005.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. ”Maximum-likelihood from incomplete data via the EM algorithm”. *J. Roy. Stat. Soc.*, 39(1):pp. 1–38, January 1977.

[7] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, USA, 2003.

[8] F. R. Kschischang, B. J. Frey and H.-A. Loeliger. ”Factor graphs and the sum-product algorithm”. *IEEE Trans. on Inform. Theory*, 47:pp. 498–519, February 2001.

[9] J.S. Yedidia and W.T. Freeman and Y. Weiss. ”Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms”. *IEEE Trans. on Inform. Theory*, 51(7), 2005.

[10] C. Herzet. Inference on graph: Study of the iterative pseudo lieklihood maximization algorithm. Technical report, available at http://www.tele.ucl.ac.be/digicom/herzet/index.php.

[11] Tom Heskes, Onno Zoeter, and Wim Wiegerinck. Approximate expectation maximization. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[12] R. Salakhutdinov, S. Roweis, and Z. Ghahramani. On the convergence of bound optimization algorithms. URL: citeseer.ist.psu.edu/584732.html.